

2026 IEEE International Conference on Software Testing, Verification and Validation (ICST 2026)

**Daejeon, Republic of Korea
18-22 May 2026**

IEEE Catalog Number:
ISBN:

CFP26TVV-POD
979-8-3195-3309-8



Copyright © 2026, IEEE

All Rights Reserved

Copyright and Reprint Permissions:

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

*** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.

IEEE Catalog Number:	CFP26TVV-POD
ISBN (Print-On-Demand):	979-8-3195-3309-8
ISBN (Online):	979-8-3195-3308-1
ISSN:	2159-4848

Additional Copies of This Publication Are Available From:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA

Phone:	(845) 758-0400
Fax:	(845) 758-2633
E-mail:	curran@proceedings.com
Web:	www.proceedings.com

TABLE OF CONTENTS

A Multimodal NLP Framework for Detecting and Removing with Explanation Code Vulnerabilities	1
<i>Jorge Guerreiro</i>	
Analyzing the Latent Input Space of LLMs for Systematic Prompt Testing	4
<i>Somin Kim</i>	
An Activation Property-Based Framework for Efficient Neural Network Verification	7
<i>Seunghyun Chae</i>	
Understanding and Overcoming Limitations of Automated Test Generation	10
<i>Lukas Krodringer</i>	
Designing and Verifying Agentic AI Systems through Structural Architecture and State Analysis	13
<i>Juyeon Yoon</i>	
Topographical Deep Learning Testing	16
<i>Gianmarco De Vita</i>	
Reinforcement Learning-Based Collaborative Fuzzing	19
<i>Hyeonmin Mo</i>	
Automated Inference of Expressive Metamorphic Relations for Effective Metamorphic Testing	22
<i>Agustín Nolasco</i>	
Improving the Accuracy of Dynamic Specification Inference	25
<i>Agustin Balestra</i>	
Dynasto: Validity-Aware Dynamic–Static Parameter Optimization for Autonomous Driving Testing	28
<i>Dmytro Humeniuk, Mohammad Hamdaqa, Housseem Ben Braiek, Amel Bennaceur, Foutse Khomh</i>	
Natural Adversaries: Fuzzing Autonomous Vehicles with Realistic Roadside Object Placements	40
<i>Yang Sun, Haoyu Wang, Christopher M. Poskitt, Jun Sun</i>	
Metamorphic Testing of Vision–Language Action–Enabled Robots	52
<i>Pablo Valle, Sergio Segura, Shaukat Ali, Aitor Arrieta</i>	
Assessing Vision–Language Models for Perception in Autonomous Underwater Robotic Software	64
<i>Muhammad Yousaf, Aitor Arrieta, Shaukat Ali, Paolo Arcaini, Shuai Wang</i>	
DeepNaqqal: Human-Aligned Automated Validation of Test Inputs for Deep Learning	76
<i>Maryam Maryam, Matteo Biagiola, Paolo Tonella, Vincenzo Riccio</i>	
How Effective Is Coverage-Guided Fuzzing to Test Deep Learning Library APIs?	88
<i>Feiran Alex Qin, M. M. Abid Naziri, Hengyu Ai, Saikat Dutta, Marcelo d’Amorim</i>	
Testing Literacy for the Era of Self-Software Systems: Rethinking Software Testing Education for AI-Mediated Development	100
<i>Nuno Pombo</i>	
A Systematic Evaluation of Environmental Flakiness in JavaScript Tests	105
<i>Negar Hashemi, Amjed Tahir, August Shi, Shawn Rasheed, Rachel Blagojevic</i>	

Pinpointing Flakiness in Web Tests via DOM Event Analysis and LLM-based Reasoning	117
<i>Yu Pei, Jeongju Sohn, Sarra Habchi, Mike Papadakis</i>	
Automatically Resolving Mismatched Invocations on Test Doubles in Test Suites	129
<i>Mengzhen Li, Mattia Fazzini</i>	
Towards an Actionable Flakiness Score	141
<i>Joanna Kisaakye, Mutlu Beyazit, Serge Demeyer</i>	
Dynamic Mutation Scheduling: Highly Parallel, Efficient Evaluation of Mutations for Rust Programs through Program Splitting	146
<i>Zalán Lévai, Donghwan Shin, Phil McMinn</i>	
How Effective are Coverage- and Diversity-Based Test Selection at Killing Stubborn Mutants?	158
<i>Islam T. Elgendy, Robert M. Hierons, Phil McMinn</i>	
Combining Input Constraints with Execution Goals	169
<i>Leon Bettscheider, Marius Smytzek, Andreas Zeller</i>	
Industrial Application of Deep Learning based Fault Localization with Mutation Features	180
<i>Heechan Yang, Ahcheong Lee, Kyutae Cho, Yunsam Kim</i>	
Constraint-Driven Fuzzing at Scale with FANDANGO	191
<i>José Antonio Zamudio Amaya, Marius Smytzek, Alexander Liggesmeyer, Valentin Huber, Andreas Zeller</i>	
TriAD: a Tricorder-based tool for detecting anomalies in software	196
<i>Caio Guimaraes Herrera, Simone R. S. Souza, Ricardo Ribeiro Dos Santos, Paulo Sergio Lopes de Souza</i>	
RIVER: An eBPF-based Runtime Verification Platform for Cyber-Physical Systems	201
<i>Dario Facchinetti, Matthew Rossi, Zhenya Zhang, Stefano Paraboschi, Paolo Arcaini</i>	
BESSER-TestGen: Automated Model-based Unit Test Generation for Low-Code Applications	206
<i>Fitash Ul Haq, Jordi Cabot</i>	
Observability and Fault Injection for LLM-Based Multi-Agent Systems in Software Engineering	211
<i>Zahra Seyedghorban, Egor Klimov, Arie van Deursen, Annibale Panichella, Burcu Kulahcioglu Ozkan</i>	
mutest-rs: Flexible, Efficient Mutation Analysis Tool for Rust Programs, using Extensive Static Analysis	216
<i>Zalán Lévai, Donghwan Shin, Phil McMinn</i>	
Consistency Meets Verification: Enhancing Test Generation Quality in Large Language Models Without Ground-Truth Solutions	221
<i>Hamed Taherkhani, Alireza DaghighFarsoodeh, Mohammad Chowdhury, Hung Viet Pham, Hadi Hemmati</i>	
How well LLM-based test generation techniques perform with newer LLM versions?	233
<i>Michael Konstantinou, Renzo Degiovanni, Mike Papadakis</i>	
Improving Automated Patch Correctness Assessment by Designing LLM-Based Oracles	244
<i>Inyeong Jang, Jinyoung Kim</i>	
Developer vs. DSpot vs. ChatGPT: A Comparative Study of JUnit Test Amplification	256
<i>David Onyango Owuor, Ajay Kumar Jha</i>	
Understanding and Improving ML-based Static Analysis Result Classification via Explainable AI	261
<i>Sai Yerramreddy, Mohammad Rafieian, Shiyi Wei, Adam Porter</i>	

Adaptive Mixing of Embeddings from Multiple Code Language Models for Fault Localization	273
<i>Juyoung Yang, EunChan Park, In-Young Ko</i>	
Assessing the Impact of Code Changes on the Fault Localizability of Large Language Models	285
<i>Sabaat Haroon, Ahmad Faraz Khan, Ahmad Humayun, Waris Gill, Abdul Haddi Amjad, Ali R. Butt</i>	
<i>Taha Khan, Muhammad Ali Gulzar</i>	
When Less Is More: Monolingual Fine-Tuning of Language Models for Industrial C# Code Review	297
<i>Igli Begolli, Meltem Aksoy, Daniel Neider</i>	
DeepFix: Debugging and Fixing Machine Learning Workflow using Agentic AI	309
<i>Fadel Mamar Seydou, Arnab Sharma</i>	
Ising-based Test Optimization and Benchmarking	314
<i>Yige Yang, Man Zhang, Tao Yue</i>	
TopoMap: Automated Topographical Mapping of the Deep Learning Input Space	319
<i>Gianmarco De Vita, Nargiz Humbatova, Paolo Tonella</i>	
Poster: NeuroCERT: Executable Testing for Oracle-Limited Biophysical Network Simulation	324
<i>Seungyoun Lee, Hyun Jae Jang, Jae-Ho Han</i>	
Poster: Transformer-Based Detection of Code Obfuscation Techniques Using Opcode Sequences	328
<i>Inwoo Jo, Wanju Kim, Eun-Sun Cho</i>	
Poster: Comparative Study of Human and Machine level prompts for LLM driven software testing	332
<i>Anand Sharma, Vivek Yelleti, Sangharatna Godbole, P. Radha Krishna</i>	
Poster: Automated Generation of a Broad Spectrum of MBA Expressions for Robust Deobfuscation Analysis	336
<i>Seoksu Lee, Sangjun An, Eunbi Cho, Eun-Sun Cho</i>	
Leveraging Large Language Models for Trustworthiness Assessment of Web Applications	340
<i>Oleksandr Yarotskyi, José D' Abruzzo Pereira, João R. Campos</i>	
Towards Automated Page Object Generation for Web Testing using Large Language Models	352
<i>Betül Karagöz, Filippo Ricca, Matteo Biagiola, Andrea Stocco</i>	
Neural Embeddings for Web Testing	364
<i>Kasun Kanaththage, Luigi Libero Lucio Starace, Matteo Biagiola, Paolo Tonella, Andrea Stocco</i>	
Automated Reproduction of Android Application Bugs with LLMs: Are We There Yet?	376
<i>Dennis Carey, Karim Elish, Paniz Abedin</i>	
GRANDSLAM: Linearly Scalable Model Synthesis	381
<i>Alexander Boll</i>	
Improving Dynamic Specification Inference with LLM-Generated Counterexamples	393
<i>Agustín Balestra, Agustín Nolasco, Facundo Molina, Diego Garbervetsky, Renzo Degiovanni</i>	
<i>Nazareno Aguirre</i>	
Systematic API Testing Through Model Checking and Executable Contracts	405
<i>Ana Catarina Ribeiro, Margarida Mamede, Carla Ferreira</i>	

Real-World Fault Detection for C-Extended Python Projects with Automated Unit Test Generation	417
<i>Lucas Berg, Lukas Krodinger, Stephan Lukasczyk, Annibale Panichella, Gordon Fraser, Wim Vanhoof Xavier Devroey</i>	
On the Evolution of Python Test Cases into Property-based Tests	429
<i>Cindy Wauters, Ruben Opdebeeck, Coen De Roover</i>	
Embedding-based Diversity Mapping for Test Generator Selection and Input Prioritization in Grammar-based Testing	441
<i>Shifat Sahariar Bhuiyan, Nargiz Humbatova, Paolo Tonella</i>	
Automated REST API Black-box Test Generation in Practice: An Experience Report from Industry	453
<i>Davide Corradini, Seung Yeob Shin, Domenico Bianculli</i>	
In Perfect Harmony: Orchestrating Causality in Actor-Based Systems	464
<i>Vladislav Mikytiy, Bernardo Toninho, Carla Ferreira</i>	
Isolating Feature Transition Errors in Dynamically Adaptive Systems	475
<i>Pierre Martou, Benoît Duhoux, Kim Mens</i>	
Synthesizing Precise Protocol Specs from Natural Language for Effective Test Generation	486
<i>Kuangxiangzi Liu, Alexander Liggesmeyer, Dhiman Chakraborty, Andreas Zeller</i>	
MT4DT: Metamorphic Testing for Digital Twins	498
<i>Philipp Zech, Sascha Hammes, Manuel Núñez</i>	
ICST Tool Competition 2026 - SDC Testing Track	503
<i>Prakash Aryan, Christian Birchler, Tommaso Fulcini, Luigi Libero Lucio, Sebastiano Panichella</i>	
RTE4SDC at the ICST 2026 Tool Competition —Self-Driving Car Testing Track	507
<i>Marcello Maugeri, PingChen Lin, Aryan Pasikhani</i>	
RoadFury at the ICST 2026 Tool Competition –Self-Driving Car Testing Track	509
<i>Chi-Nguyen Tran, Dao Sy Duy Minh, Huynh Trung Kiet, Phu-Hoa Pham, Nguyen Lam Phu Quy</i>	
KSERESNET at the ICST 2026 Tool Competition – Self-Driving Car Testing Track	511
<i>Vishal Kumar Swain, Sangharatna Godbole, P. Radha Krishna, Avijit Das</i>	
ICST Tool Competition 2026 – UAV Testing Track	513
<i>Erdem Uysal, Gregory Loubet-Bonino, Ali Javadi, Aryan Prakash, Aren A. Babikian, Dmytro Humeniuk Sajad Khatiri, Sebastiano Panichella</i>	
TLQD at the ICST 2026 Tool Competition – UAV Testing Track	517
<i>Hainan Zhang, Shixuan Zhou, Haoxiang Qin, Yi Xiang</i>	
Do Language Models Prefer Vulnerable Code? A Probabilistic Study of Insecure Code Preference	519
<i>Rui Melo, Sofia Reis, Andre Catarino, Rui Abreu</i>	
RAVEN: Mining Defensive Patterns in Ethereum via Semantic Transaction Revert Invariants Categories	531
<i>Mojtaba Eshghie, Melissa Mazura, Alexandre Bartel</i>	
Individual Fairness Testing in Fairness through Unawareness	543
<i>Taisei Kuma, Takashi Kitamura, Shingo Takada</i>	

Formalizing Test Levels for Architectural Alignment and Overall Optimization	555
<i>Riku Ogura, Ren Karita</i>	
FALCON: Efficient Test Case Prioritization via Submodular Optimization	566
<i>Twumasi Mensah-Boateng, Jing Yuan, Hyunsook Do</i>	
DANTE: Data-Driven Test Case Selection and Prioritization for Long-Running Test Suites	577
<i>Simone Reale, Elisabetta Di Nitto, Luciano Baresi, Massimiliano Di Penta, Giovanni Quattrocchi</i>	
How Effectively Do Large Language Models Help with Build Conflict Resolution?	589
<i>Sheikh Shadab Towqir, Xiaoyan Zang, Todd Mytkowicz, Na Meng</i>	
Taming the Variants Multi-Architecture Continuous Testing at Google	601
<i>Tim A. D. Henderson, Sushmita Azad, Chandrakanth Chittappa, Ali Esmaeeli, Laura Macaddino</i> <i>Sam Manfreda, David Margolin, Dharma Naidu, Sabuj Pattanayek, Sachin Sable, Ruslan Sakevych</i> <i>Dushyant Acharya, Adrian Berding, Kevin Crossan, Wolff Dobson, Avi Kondareddy, Abhayendra Singh</i>	
Continuous Requirement Conformance Falsification for Cyber-Physical Systems via Cross-Revision Knowledge Transfer	612
<i>Jesper Winsten, Ivan Porres</i>	
Towards Counterfactual Explanation and Assertion Inference for CPS Debugging	623
<i>Zaid Ghazal, Hadiza Yusuf, Khoulood Gaaloul</i>	
On the Freshness of Pinned Dependencies in Maven	635
<i>Vasudev Vikram, Yuvraj Agarwal, Rohan Padhye</i>	
Android Instrumentation Testing in Continuous Integration: Practices, Patterns, and Performance	647
<i>Hamid Parsazadeh, Taher A. Ghaleb, Safwat Hassan</i>	
Predictive Pull Request Batching to Accelerate Merge Pipelines in Continuous Integration at Scale	659
<i>Maximilian Jungwirth, Martin Gruber, Gordon Fraser</i>	
Intelligent resource prediction for SAP HANA continuous integration build workloads	671
<i>Torsten Mandel, Jonathan Bader, Hanyoung Yoo, Stephan Kraft</i>	
Lost in Navigation: Detecting Keyboard Navigation Accessibility Issues in Web Pages	682
<i>Robert Winn, Paul T. Chiou, William G. J. Halfond</i>	
SemLink: A Semantic-Aware Automated Test Oracle for Hyperlink Verification using Siamese Sentence-BERT	694
<i>Guan-Yan Yang, Wei-Ling Wen, Shu-Yuan Ku, Farn Wang, Kuo-Hui Yeh</i>	
A Linguistically Motivated Automated Testing Framework For ASR Accent-Robustness	706
<i>Margot Masson, Thomas Laurent, Anthony Ventresque</i>	
LLM2Alloy: Investigating LLM-Generated Formal Specifications for Automated Test Derivation in Production Software	717
<i>Tasmim Rashid, Muhammad Zubair Malik</i>	
An Adaptive Hybrid Quantum-Classical Framework for Test Suite Minimization via Quantum Approximate Optimization Algorithm	722
<i>Laksh Sharma, Krish Sharma, Ajay Kumar</i>	

XPJudge: Explanation-Aware Pairwise Judging for Cost-Effective Code LLM Selection	732
<i>Yasmine Ben Braiek, Housseem Ben Braiek, Foutse Khomh</i>	
Quantum Circuit Repair by Gate Prioritisation	744
<i>Eñaut Mendiluze Usandizaga, Thomas Laurent, Paolo Arcaini, Shaukat Ali</i>	
Toward Live Noise Fingerprinting for Discrepancy Analysis in Quantum Software Engineering	749
<i>Avner Bensoussan, Elena Chachkarova, Karine Even-Mendoza, Sophie Fortz, Vasileios Klimis</i>	
<i>Mohammad Reza Mousavi</i>	
Keynote Talk: Adversarial Benchmarking An Urgent Call for a Stronger Signal of Technological Progress	755
<i>Marcel Böhme</i>	